DOCUMENT RESUME

ED 279 702 TM 870 089

AUTHOR Venezky, Richard L.
TITLE Literacy and the NAEP Reading Assessments. A Background and Position Paper Prepared for the NAEP Reading Review Subcommittee.
PUB DATE Aug 86
NOTE 22p.; One of 46 papers commissioned by the Study Group on the National Assessment of Student Achievement and cited in Appendix B to their final report "The Nation's Report Card" (TM 870 049). For other papers in this group, see TM 870 050-094.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adult Literacy; Cognitive Processes; *Educational Assessment; Elementary Secondary Education; Functional Literacy; *Literacy; Measurement Objectives; National Surveys; *Reading Skills; Reading Tests; *Scaling; *Scoring; *Testing; Testing Programs; Test Results
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

This paper reviews literacy and literacy assessment in America, in relation to the various National Assessment of Educational Progress (NAEP) reading assessments. Literacy can be defined and assessed only in relation to literacy demands. A current concept of literacy distinguishes between reading comprehension as an outcome of lexical access and comprehension processes, and literacy as the application of particular skills in a social context. Closely related to the changing definitions of literacy are the techniques employed for assessing literacy. The most refined literacy assessment to date is the NAEP Young Adult Literacy Survey. Analysis of the responses from this study show multiple dimensions to literacy. NAEP reading objectives have shifted from a behavioral to a process-oriented view. Scale scores (which may reduce the educational value of a test, but offer psychometric convenience) have replaced estimated percentage correct for items or item types. Several recommendations are made: (1) the NAEP Reading Assessment should focus on surveying the skills and attitudes of real-world literacy in youth and young adults; (2) NAEP's Reading Assessment should cover high school dropouts and young adults, as well as in-school students; (3) public reporting of NAEP assessment scores should emphasize item types, using weighted percentage correct; and (4) NAEP should report results objectively, without pressure to promote political agenda. (LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED279702

Literacy and the NAEP Reading Assessments

Richard L. Venezky

University of Delaware

Paper Commissioned by

THE STUDY GROUP ON THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT

1986

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. L. Venezky

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

TM 870 089

Literacy and the NAEP Reading Assessments

A Background and Position Paper
Prepared for the NAEP
Reading Review Subcommittee

by

Richard L. Venezky
University of Delaware
August 1986

Introduction

Literacy is a set of abilities, an attitude, an everyday necessity, and a national resource. It is the primary tool of adaptation and self-improvement in a rapidly changing technological society, yet those weakest in literacy skills are also the ones least capable of directing their own betterment and the least capable of assisting their children in obtaining a better share of this country's resources and opportunities. By most rational standards, America has unusually low illiteracy. America has not had a pressing illiteracy problem for many years, at least not as it had 150 years ago, but by some definitions it has a literacy problem.

Many young adults have generally low literacy abilities relative to the demands of employment, citizenship, and individual advancement. Furthermore, since the 1970s when detailed surveys of literacy skills were first undertaken in the USA, America's minorities and its poor have scored disproportionately lower than middle class whites, even when respondents' education and other background characteristics are controlled. A recent survey of literacy skills in young adults, 21-25 years of age, shows that this problem is still present in approximately the same magnitude as 15 years ago. This paper is a review of literacy and literacy assessment in America, discussed in relation to the various NAEP reading assessments. From this

discussion several conclusions are drawn concerning how the NAEP reading tests could contribute to an understanding of America's literacy problems.

Defining Literacy

Literacy is an ability that can be defined and assessed only in relation to literacy demands. By definition it is both functional and relative. In the Middle Ages when Latin was the *lingua franca* of religion and law, a *litteratus* was one learned in Latin, not one with a rudimentary ability to read and write (Clanchy, 1983, p.16). Vernacular literacy developed with the increased bureaucracy of secular governments, beginning roughly in the 12th century, and was extended towards the masses by the Reformation and succeeding religious and political upheavals.

Concepts of literacy changed slowly in the first 200 years of America's settlement. Initially, rudimentary ability to read aloud was considered sufficient for the average person, with writing ability added at a later time. By the middle of the 19th century understanding was an essential component of competent reading. When the Census Bureau began in 1840 to include a literacy question in its household surveys, no criterion for literacy ability was offered. Respondents were asked if they could read and write and a simple yes or no was recorded. (In time the question was refined to "Can you read and write a simple message?", but in recent years this has generally been asked only of those with less than five or six years of schooling. All others were asssumed literate (Bureau of the Census, 1971).)

By the 1880s a more demanding sense of literacy was exhibited, especially by the courts. For example, in 1886 the Texas Court of Appeals addressed the issue of whether or not signing one's name was sufficient for meeting the state literacy requirement for jury duty. The Appeals Court in this case ruled no, claiming

> "That a person can write his name certainly does not fill the measure of the statutory requirement that the juror should be able to *write*. We think that he should be able to express his ideas in words upon paper with pen or pencil." (21 Texas Ct. App. 379; cited in Stevens, 1983. p.xx).

Testing in World War I of 1.522 million soldiers and sailors included tests of "ability to read and understand newspapers and to write letters." (Gray and Munroe, 1929, pp.21-22). In the 1950s UNESCO, while portraying literacy as a continuum of skills, suggested two levels of literacy for international comparisons: a minimal level, marked by the ability to read and write a simple message, and a functional level, demonstrated by literacy sufficiently high for the person to function in a social setting (UNESCO, 1957, p.18). The definition of functional literacy given by UNESCO appears to derive from Gray's somewhat circular definition:

> "A person is a functional literate when he has acquired the knowledge and skills in reading and writing which enable him to engage effectively in all those activities in which literacy is normally assumed in his culture or group." (Gray, 1956, p.19).

In the 1970s several national assessments of adult knowledge and ability established a broader view of literacy as a complex of skills centered on the print-related demands of everyday life, including work, civic duties, and recreation. For example, the Adult Performance Level Study (Nafzinger et al., 1975) defined four basic areas of skill proficiency: communication, computation, problem solving, and interpersonal skills. These were tested across five knowledge areas: occupational, consumer economics, community resources, government and law, and health. Similarly the U.S. Office of Education in 1974 defined the following eleven skills for functional literacy.

(1) Read and understand all sections of a newspaper.

(2) Read driver's license test.

(3) Read voter registration instructions.

(4) Read popular business contracts (car dealers, furniture stores).

(5) Read labels on groceries, medicines, etc.

(6) Read entry level job materials.

(7) Read personal letters, bills.

(8) Read public instructions-road and building signs.

(9) Read telephone directory.

(10) Read job application forms.

(11) Read business letters from debtors and creditors.

More narrow definitions of literacy, but still centering on real-life print-related tasks, were offered by Louis Harris & Associates (1970) and the National Center for Health Statistics (Vogt, 1973).

A NAEP survey of adult work skills and knowledge carried out in 1973-74, and reported in 1976, assessed six different areas, the first four of which could be loosely tied to a broad concept of literacy: computation and measurement skills, graphic and reference material skills, written

communication skills, manual and perceptual skills, general job knowledge, and specific job knowledge (NAEP, 1976). Although young adults in the age range 26-35 were assessed, many of the questions used in this assessment were similar to those used in the most recent assessment of young adults, covering the age ranges 21-25.

One distinction that is important for deriving a current concept of literacy is the difference between reading comprehension and literacy. The former is typically defined as an outcome of two cognitive processes, lexical access and comprehension (e.g., Perfetti, 1985, p.4). Literacy on the other hand involves the application of particular skills in a social context. Nafzinger et al., (1975, p.19), for example, claim "literacy, in contrast to reading, implies both reading skills and socially appropriate reading behavior..." Literacy implies reading and writing for a purpose generally related to social utility or individual advancement. Sticht (1975) suggests that the definition of literacy exclude reading for pleasure. If we accept this, then we have an opposition between the schol-based reading curriculum, which is strongly oriented towards reading fiction for personal pleasure and the needs of literacy for social and individual advancement. The former tends towards elitism, stressing the correct processing of acceptable texts; the latter is democratic, stressing ability for social and individual utility.

It is interesting to note that in the *Oxford English Dictionary* the term *illiteracy* is evidenced at least as early as 1660 and has apparently been in continuous use since then. The term *literacy*, on the other hand, is cited first in 1883 in the *New England Journal of Education*. This disparity in dates of introduction of the two terms reflects a change from the earlier period in this country's history when illiteracy and its erradication was the main focus of society, to around the end of the 19th century when the quality of literacy

itself became an issue.

Literacy Assessment

Closely related to the changing definitions of literacy over the past several hundred years are the techniques employed for assessing literacy. In medieval times, a person guilty of a serious crime avoided the death penalty by proving himself a *litteratus*. This usually required reading aloud from a Psalter or some other type of prayer book (Gabel, 1929). This approach to reading assessment remained until the second decade of the 20th c., even though written tests were begun in other subjects as early as 1864 (Calhoun, 1973). Even Joseph Mayer Rice, who in his school surveys used written tests in other subjects, depended upon subjective evaluations of students reading aloud for assessment of reading ability (Rice, 1892).

The earliest reading tests (e.g., Starch, 1915) set the pattern for most of the school-based comprehension tests of today: passages of one or two paragraphs, followed by one or more multiple-choice questions. Generally, vocabulary and comprehension have been tested, along with word analysis skills at the lower levels. Use of real-life documents and reading tasks, as opposed to narrative or expository passages, was popularized in the 1970s, through the Louis Harris & Associates survey (1970), which required filling out different forms, the APL Survey, which drew on a wide range of different document types (Nafziger et al., 1975), and the NAEP *Adult Work Skills and Knowledge Assessment* (NAEP, 1976).

The most refined adult literacy assessment to date is the *Young Adult Literacy Survey* which NAEP will release this September, (Kirsch & Jungeblut, 1986). In this study, mostly open-ended questions were used, simulating tasks drawn from everyday print experience. Respondents were asked to locate information in an almanac; fill in phone messages, job

applications, check ledgers, and other forms; read a passage aloud; explain orally how to figure interest on a bank loan advertised in a newspaper; etc. Tasks were varied to present a range of processing demands and response forms. The analyses of the responses from this study show multiple dimensions to literacy, as opposed to the single scale reported in the last NAEP Reading Assessment (NAEP, 1985).

NAEP Reading Objectives

From the earliest NAEP Reading Assessment, NAEP reading objectives were formulated to satisfy three constituencies: scholars in the field, educators (i.e., teachers), and thoughtful lay citizens. However, the need to orient objectives toward real life tasks was strongly stressed in the first reading assessment. "Parents and others interested in education should agree that an objective...is of value in modern life" (NAEP, 1970, p.2). For the first reading assessment, two firms, ETS and SRA, were invited to submit objectives. SRA's were selected, but with requests for modifications. The resulting objectives, which are listed below, were to be applied to four types of reading materials for generating test items: literature, academic and expository prose, utilitarian writing (forms, labels, job-related materials, etc.), and reference materials. The objectives were described in behavioral terms:

Comprehend what is read (including decoding and vocabulary)

Analyze what is read (sequence, structure, author's craft)

Use what is read (remember significant facts, follow directions, etc.)

Reason logically from what is read (inferences, deductions, etc.)

Make judgments from what is read (validity of facts, detecting false logic, etc.)

Have attitudes about and interest in reading (is willing to read, etc.)

However, when the results of the initial assessment were analyzed, the NAEP staff decided to report outcomes under nine themes:

(1) Word meaning

(2) Visual aids (drawings and pictures; signs and labels; charts, maps, graphs; forms)

(3) Written directions

(4) Reference materials

(5) Significant facts in passages

(7) Inferences from passages

(8) Critical reading of passages

(9) Reading rate

These themes appear to reflect three main categories of literacy information; viz., *basics of reading* (word meanings, reading rate), *everyday literacy tasks* (visual aids, written directions, and reference materials), and *school-based reading tasks* (significant facts, main idea, inferences, critical reading). They were, nevertheless, cumbersome to use for evaluation or policy making.

For the projected 1973-74 reading assessment, a switch was made to in-house development of objectives, with outside contracting of assessment items (exercises). After several false starts, numerous recyclings, and a one-year delay in the assessment itself, a new set of behavioral objectives were accepted. These were:

I. Demonstrates behavior conducive to reading.

II. Demonstrates word identification skills.

III. Possesses skills for reading comprehension.

IV. Uses a variety of approaches in gathering information.

The objectives were viewed in the spirit of their time as defining "various

steps in the reading process." Tasks based on these objectives were assumed to "encompass a hierarchy of knowledge and skills ranging from elementary to complex" (NAEP, 1974, p.6).

However well-intended the assessment design was in relationship to NAEP's mission and constituencies, the results of the 1974-75 assessment were released through a panel of university-based reading specialists, who reorganized the results in traditional academic categories of literal and inferential comprehension, and reference skills (NAEP, 1976). Thus a survey designed to reveal a broad range of information on reading was narrowly focused on the elementary school curriculum for reporting.

For the 1979-80 assessment, funding cuts forced NAEP to combine the reading and literature assessments, resulting in objectives and test items that were strongly biased toward the school reading/literature curriculum and away from literacy in the broad sense. Decoding skills were excluded, and almost all exercises not centered on narrative or expository paragraphs were dumped into subobjectives in the study skills category ("Obtains Information from Nonprose Reading Facilitators," "Uses the Various Parts of a Book," "Obtains Information From Materials Commonly Found in Libraries or Resource Centers.")

Ideas current in reading and literary research were incorporated into the objectives, including the view of reading as an interactive process and the focusing of meaning on propositions. The authors claimed that "meaning is not in the print alone," that the "reader brings his or her knowledge and experience to bear in the reading task" (NAEP, 1980, p.2). Yet to distinguish "responding to written works" from comprehension, the latter was limited "to the largely unconscious activity of understanding primary meaning in the process of reading" (p.2).

The last major overhaul of the reading objectives was done for the 1985-86 reading assessment, and is not yet published. A draft document describes four categories of objectives, three defined by text processing concepts and a fourth for background ;information on reading habits and attitudes. These are: Deriving Information; Integrating and Applying Information; Evaluating and Reacting; and Reading Habits, Preferences, and Values. Test items were to be defined for each of the first three objectives using a full range of text types: stories, textbooks, encyclopedias, magazine articles, business letters, order forms, maps, advertisements, graphs, etc. While it has not been possible to examine the test items to see how closely they follow this design, the test-type-by-processing-demands approach reflected here closely parallels the design for the Young Adult Literacy Survey just completed by NAEP.

Surveys of reading habits, preferences, and values add an important dimension to literacy asssessment that most other surveys have ignored. Gray and Munroe, in a landmark study of adult reading habits, emphasized the importance of attitude toward reading as revealed by reading preferences (1929, p.264).

> "Many adults fail to read...because they are not interested...Of even greater significance than the amount read by adults is the character or quality of the material that is read."

The recent study of young adults mentioned above not only gathered extensive data on reading practices, but linked these data to reading ability, using multivariate techniques. The NAEP Reading Assessments, in contrast,

have gathered background data on reading practices, but have made only small amounts of it publicly available.

In summary, one can trace a shift in the NAEP Reading Objectives from the behavioral view of reading that was prevalent in the 1960s to a process-oriented view for the 1980s, with a continuing attempt to mix a curriculum-based view of reading with a view that more closely approximates the literacy demands of everyday life. Test items, however, continue to be multiple choice, based mostly on paragraph reading. This limitation leaves a large gap between assessment tasks and the true tasks of everyday literacy, in that the latter require in most instances creation of a response rather than selection from alternatives. Multiple choice items are reasonable proxies for free-response abilities, but they fail to create the appropriate model for educators and policy makers as to the tasks that students and young adults can or cannot do. That is, multiple-choice items may have external validity, but they lack face value.

Scale Scores

One of the major changes which ETS has instituted since taking over the NAEP is to focus its reporting of reading scores not on estimated (i.e., weighted) percentages correct for items or item types, but on scale scores. These latter entities are derived through non-order preserving transformations of the weighted item scores (and of certain item characteristics). Scale scores eliminate some of the problems inherent in item scores, and also make linkage across tests easier, but they also reduce the educational value of the test.

For example, scale scores lose some of their psychometric value if more than one scale exists. With more than one scale, comparison of item values

for items on different scales is not valid. This, in part, may explain why an assessment designed to probe different dimensions of reading was reduced to a single scale for reporting. (But NAEP, nevertheless, was forced to use mean percentages correct for comparing scores on different types of questions across assessments.) *This reduction to a single scale has no justification other than psychometric convenience.* It eliminates any value that the NAEP Reading Assessment might have had to understanding the dimensions of literacy.

At issue is not the ability of NAEP to generate a transformation for projecting disparate items onto a single scale; this they apparently succeeding in doing, although a number of items had to be eliminated to do it (e.g., reference skills). Rather, the issue is what information might be of value for educational decision making, and here the representation of all reading abilities as a single factor has limited value. Whether the most useful dimensions of literacy are captured by literal and inferential comprehension, plus study skills, or by some other classification remains to be determined. Nevertheless, reading instruction and reading assessments are planned according to different skill categories. By eliminating all distinctions among reading skills, NAEP has reversed almost 100 years of progress in the study of reading abilities.

A second objection to scale scores is that without any clear anchoring to standardized reading measures, their values are uninterpretable. The arbitrary creation of levels of proficiency for the 1984 reporting with 50 scale points separating successive levels, is not a useful substitute for scores anchored to either reading materials or nationally standardized proficiency levels. The current levels give a misleading impression of scientifically derived divisions, when in fact mostly subjective judgments were involved.

The search for dimensions of reading ability will not be resolved by a national assessment such as NAEP. This is a complex issue that will yield its secrets only to carefully planned research, and then probably not too quickly.

Recommendations

1. The primary focus of the NAEP Reading Assessment should be on surveying the skills and attitudes of real-world literacy in youth and young adults. An emphasis on a school-based definition of reading comprehension does not provide the information needed for educational policy. Literacy is a set of skills that extend beyond reading comprehension. Literacy skills often require integration of reading, writing, listening, and speaking, and always require application. The reading skills that schools teach may not be the skills that need to be taught to create adequate literacy, therefore, an assessment that surveys only what the schools teach may not tell whether or not literacy skills are adequately learned. The direction that the 1985-86 NAEP Reading Objectives took should be extended to match more closely the item-generating approach of the Young Adult Literacy Survey. A single survey should be developed for in-school students, dropouts, and young adults.
2. NAEP's Reading Assessment should regularly cover high school dropouts and young adults, as well as in-school students, as was done in the 1970-71 survey. To ignore these first two categories is to ignore a major part of NAEP's original mission. If the nation's information needs in education were solely on in-school students, we probably would have no use for NAEP. The same psychometric effort now exerted on I.R.T. scaling could probably create similar linkages with available data from standardized reading scores.
3. The primary emphasis in public reporting of NAEP assessment scores should be on item types, using weighted percentage correct. Regardless of the uncertainties inherent in such scores, they remain the most

interpretable descriptors for educators and policy makers. More powerful scales are of no value if they are mismatched to the interpretive powers of the Assessment audience or if they eliminate the data relationships that are critical for policy.

4. NAEP is in a delicate and often precarious position. Politicians and educators would like to see signs of improvement in each assessment, evidence that the enormous national and state investments in education have been well spent, that students are performing better and that the large disparities across racial and ethnic groups are disappearing. But invariably the news has been mixed; some things to cheer about and some things to cause concern. NAEP has subtitled its most recent reading report (*The Reading Report Card*), "Progress Toward Excellence in Our Schools." Yet a less optimistic interpretation of the report might have led to a title like "Failure to Achieve Excellence in Our Schools" or "Stumbling Blocks to Excellence in Our Schools." After all, the 1984 results show 17-year-old blacks reading no better than 13-year-old whites, and 9-year-old blacks making no gains from 1980 to 1984, suggesting that the large gains made by black 13- and 17-year-olds over the 1980-1984 interval, in particular, will taper off in 1988.

   My concern is not with which subtitle best fits the result, but with the ability of NAEP to report results objectively, without pressure to promote anyone's political agenda. Perhaps the Census Bureau is a good model for what NAEP should be. For the most part the Census Bureau reports the unadorned facts, free of value-judgments. Its publications rarely give the sense that the Census Bureau management would prefer to see one outcome rather than another. Somehow this same attitude needs to be imported to NAEP.

References

Bureau of the Census (1971). *Illiteracy in the United States: November 1969.* Current Population Reports, Series P-20, No. 217. U.S. Department of Commerce, Washington, D.C.

Calhoun, D. (1973). *The intelligence of a people.* Princeton, N.J: Princeton University Press.

Clanchy, M. (1983). Looking Back from the invention of printing. In D. Resnick (Ed.), *Literacy in historical perspective.* Washington D.C.: Library of Congress.

Gabel, L.C. (1929). Benefit of clergy in the later middle ages. *Smith College Studies in History, 14,* pp.xx-yy.

Gray, William S. (1956). *The teaching of reading and writing.* Paris: UNESCO.

Gray, William S., and Munroe, Ruth (1929). *The reading interests and habits of adults* . New York: Macmillan.

Harris, Louis, and Associates (1970). *Survival literacy.* New York: Louis Harris and Associates.

Kirsch, I.S., and Jungeblut, A. (1986). *Literacy: Profiles of America's young adults.* Report No. 16-PL-02. National Assessment of Educational Progress, ETS, Princeton, N.J.

NAEP (1970). *Reading objectives.* Ann Arbor: NAEP.

NAEP (1974). *Reading objectives: Second assessment.* Denver: NAEP.

NAEP (1976). *Adult work skills and knowledge.* Denver: NAEP.

NAEP (1980). *Reading and literature objectives: 1979-80 assessment.* Denver: NAEP.

NAEP (1985). *The reading report card: Progress toward excellence in our schools.* Princeton, N.J.: ETS.

Nafziger, D.H., Thompson, R.B., Hiscox, M.D., Owen, T.R. (1975). *Tests of*

functional adult literacy: An evaluation of currently available instruments. Portland, OR: Northwest Regional Educational Laboratory.

Perfetti, C. (1985). *Reading ability.* New York: Oxford University Press.

Rice, J.M. (1893). *The public-school system of the United States.* New York: The Century Co.

Starch, D. (1915). The measurement of efficiency in reading. *Journal of Educational Psychology, 6,* 1-24.

Stevens, E. (1983). Illiterate Americans and nineteenth-century courts: The meanings of literacy. In D. Resnick (Ed.), *Literacy in historical perspective.* Washington, D.C.: Library of Congress.

Sticht, T.G., (Ed.) (1975). *Reading for working: A functional literacy anthology.* Alexandria, VA: Human Resources Research Organization.

UNESCO (1957). *World illiteracy at mid-century.* Paris: UNESCO.

Vogt, D. (1973). Literacy among youths 12-17. *Vital and Health Statistics, Series 11-131.* Washington, D.C.: U.S. Government Printing Office.